

本周周报（2013.11.18-2013.11.24）

郭方舟

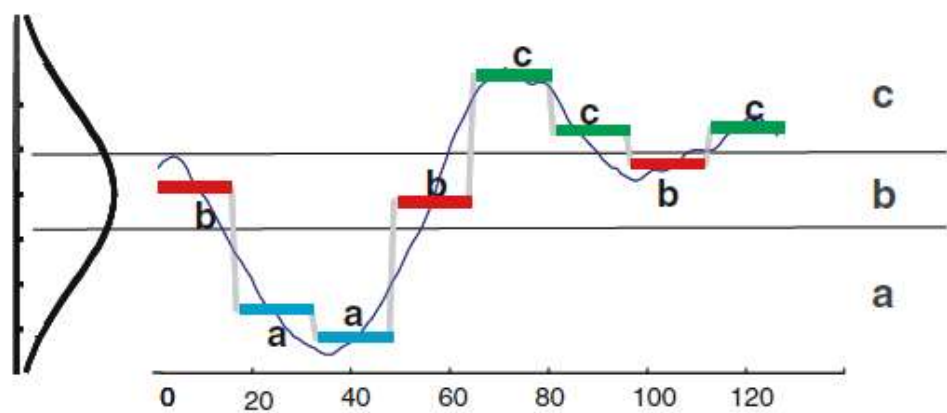
本周工作

1. 空气污染数据可视化

本周主要完成了三个工作：

1. 使用 sax 方法对数据进行处理：

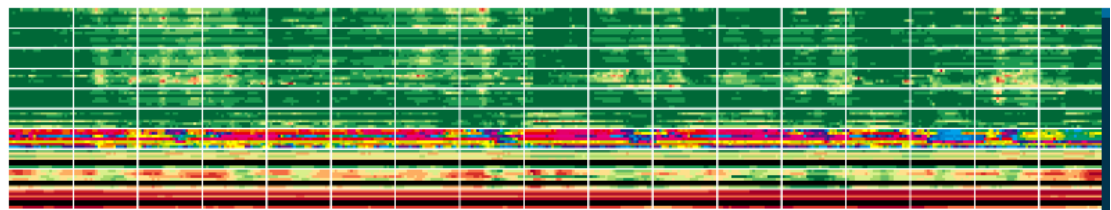
首先使用 Z-Score 标准化将原始数据标准化，使得原始数据符合标准正态分布，然后再通过分割使得每个分割区域的点数近似相等，如图图所示：



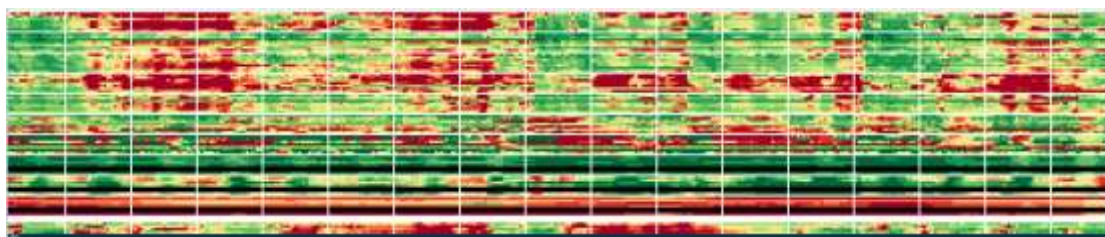
这个方法的作者也给出了参考的分割表：

β_i	a							
	3	4	5	6	7	8	9	10
β_1	-0.43	-0.67	-0.84	-0.97	-1.07	-1.15	-1.22	-1.28
β_2	0.43	0	-0.25	-0.43	-0.57	-0.67	-0.76	-0.84
β_3		0.67	0.25	0	-0.18	-0.32	-0.43	-0.52
β_4			0.84	0.43	0.18	0	-0.14	-0.25
β_5				0.97	0.57	0.32	0.14	0
β_6					1.07	0.67	0.43	0.25
β_7						1.15	0.76	0.52
β_8							1.22	0.84
β_9								1.28

原来的效果：

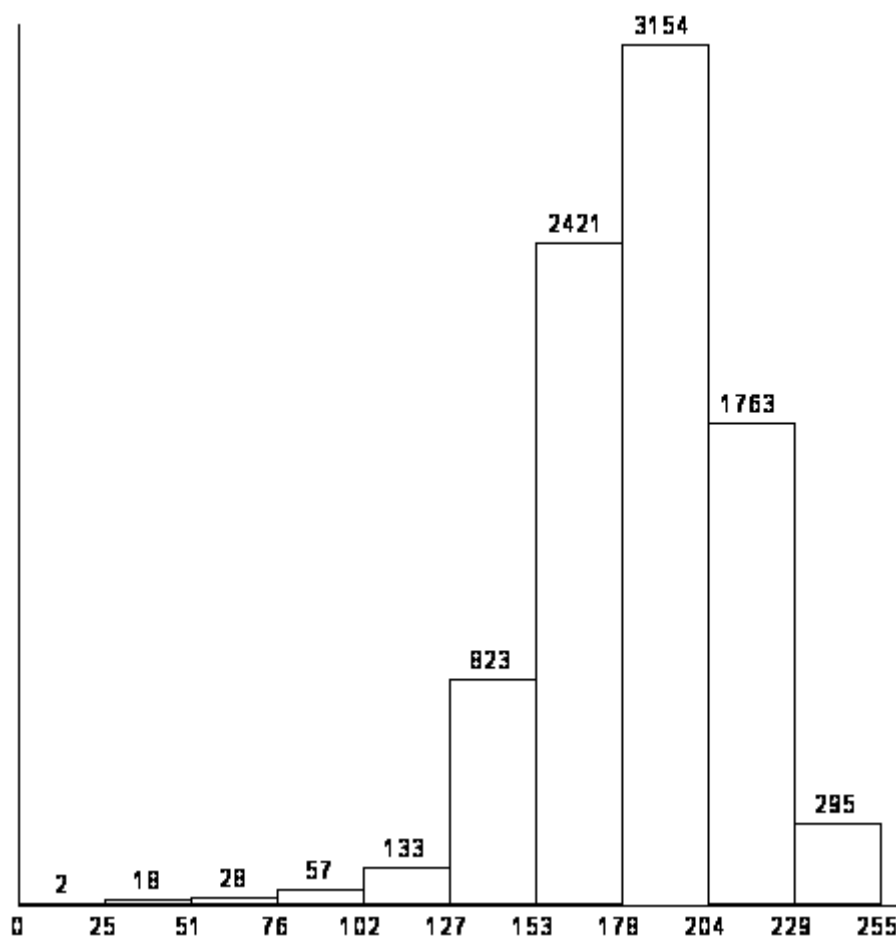


现在的效果：



可以看到现在的可视化效果中，处于高等级的点增多了，这一方面使相对高值的区域更加显眼，另一方面也使整个可视化效果鲜艳许多。

2. 对 ck 距离的结果绘制了一个直方图，来观察 ck 距离的分布；

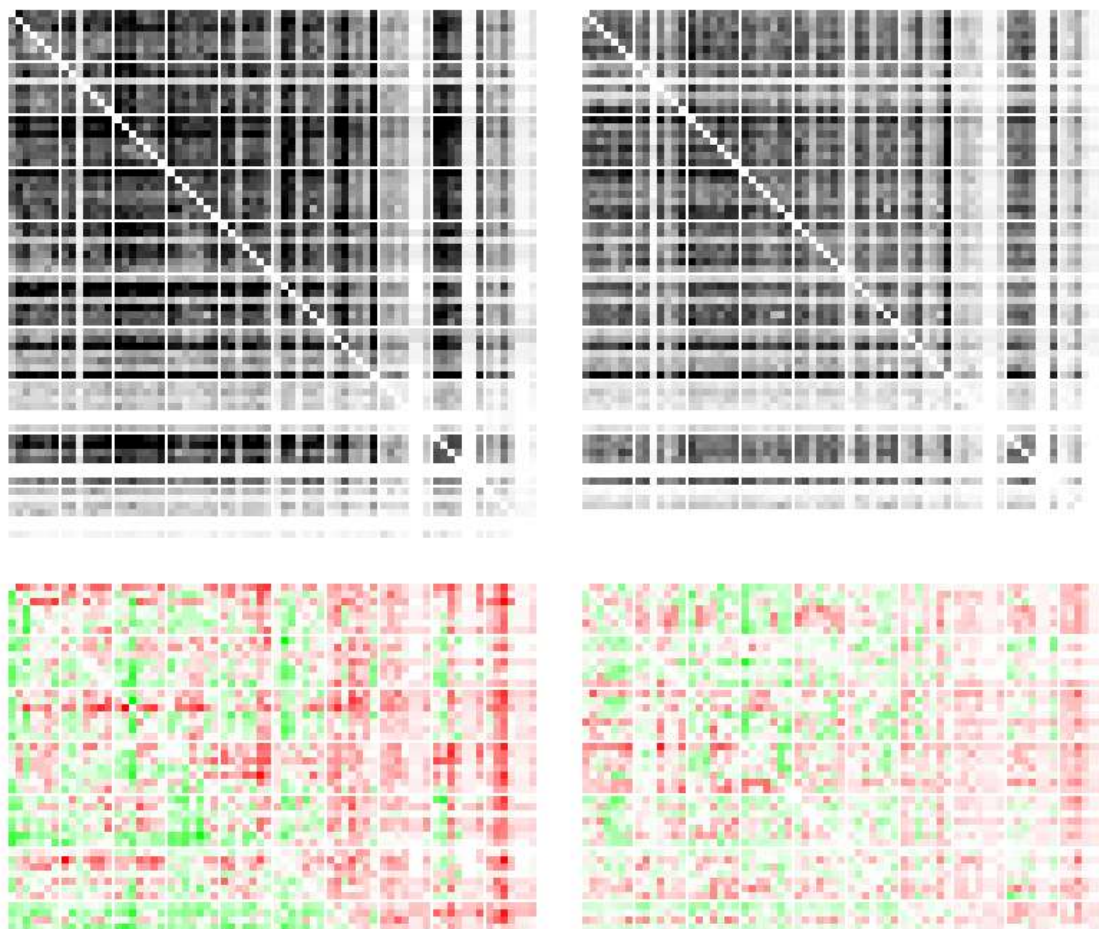


计算出来的 ck 距离已经被变换到 0-255 区间。由于 ck 距离越小代表两张图片越相似，因此在投影时将对应值进行了扭转，即最小值变换为 255，最大值变换为 0。

3. 计算转移熵

由于 multi-var 的转移熵目前没有好的方法计算，所以本周计算了单属性单传感器之间的转移熵。数据中共有 7 个传感器，每个传感器下有 10 个属性，这样单属性单传感器两两组合就有 70*70 种情况，将这些值使用矩阵的形式可视化出来。

目前的计算方案是将一年分为四个季度，每个季度计算转移熵。计算转移熵后算出 $H_{i \rightarrow j}$ 和 $H_{j \rightarrow i}$ 之间的差值，一对转移熵有两个差值，互为相反数，将这些差值也用矩阵的形式进行可视化。



由于图比较大，没有办法截取全部的矩阵，所以截取部分做展示。上面四个矩阵是转移熵，下面四个矩阵是转移熵之间的差值。差值为绿色时表示差值为负，差值为红色时表示转移熵为正。

目前从转移熵中能看到的只有空气湿度对于各空气污染属性的影响较大，印证了我们之前的猜想。

下周工作

空气污染数据的可视化

跟斐然师兄讨论，改进转移熵的计算和使用，看看能不能使用转移熵挖掘出更多有用的信息。